

STAT C206B: Topics in Stochastic Processes

Feynman Liang*
Department of Statistics, UC Berkeley

Last updated: February 4, 2020

Contents

1	Lecture 4: Daniell integration	1
2	Lecture 5: Extreme point representation of measures	5
«««< HEAD		

1 Lecture 1: Background material

2020-01-21

1.1 Ferguson distributions / Dirichlet processes

Definition 1 (*Gamma distribution*)

Random variable X supported on $(0, \infty)$ has *Gamma distribution* with shape $\alpha > 0$ and inverse scale / rate $\beta > 0$, written $X \sim \text{Gamma}(\alpha, \beta)$ if it has density

$$f_X(t) = \mathbb{1}\{t \in (0, \infty)\} \frac{\beta^\alpha t^{\alpha-1} e^{-\beta t}}{\Gamma(\alpha)} \quad (1)$$

where $\Gamma(t) = \int_0^\infty u^{t-1} e^{-u} du$ is the Gamma function defined for all $\Re t > 0$ and analytically continued to $\mathbb{C} \setminus \{n \in \mathbb{Z} : n < 0\}$

Proposition 2 (*Gamma closed under summation*)

If $Y \sim \text{Gamma}(\alpha, \beta)$ and $Z \sim \text{Gamma}(\gamma, \beta)$ are independent, then $Y + Z \sim \Gamma(\alpha + \gamma, \beta)$.

Proof.

$$\begin{aligned} f_{Y+Z}(t) &= \int_0^t f_Y(u) f_Z(t-u) du \\ &= \frac{1}{\Gamma(\alpha)\Gamma(\gamma)} \beta^{\alpha+\gamma} e^{-\beta t} \int_0^t u^{\alpha-1} (t-u)^{\gamma-1} du \\ &= \frac{1}{\Gamma(\alpha)\Gamma(\gamma)} \beta^{\alpha+\gamma} e^{-\beta t} \int_0^1 (tv)^{\alpha-1} (t-(tv))^{\gamma-1} t dv \\ &= \frac{1}{\Gamma(\alpha)\Gamma(\gamma)} \beta^{\alpha+\gamma} e^{-\beta t} t^{\alpha+\gamma-1} B(\alpha, \gamma) \end{aligned}$$

where $B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ is the beta function □

A closely related distribution obtained from concatenating Gamma random variables into a vector and then normalizing the sum to 1 is the Dirichlet distribution.

Definition 3 (*Dirichlet distribution*)

*feynman@berkeley.edu

Let $\alpha \in (0, \infty)^K$. Random (probability) vector X taking values on the $K - 1$ -dimensional probability simplex $\Delta^{K-1} = \{x \in [0, 1]^K : \sum_i x_i = 1\}$ has *Dirichlet distribution* of order K and concentration parameters α , denoted $X \sim \text{Dir}(\alpha)$, if it has density

$$f_X(x) = \mathbb{1}\{x \in \Delta\} \frac{\Gamma\left(\sum_{i=1}^K \alpha_i\right)}{\underbrace{\prod_{i=1}^K \Gamma(\alpha_i)}_{=: B(\alpha)^{-1}}} \prod_{i=1}^K x_i^{\alpha_i - 1}$$

Proposition 4 (Constructing Dirichlet from Gammas)

Let X_1, \dots, X_n be independent $\text{Gamma}(\alpha_i, \beta)$ distributed, $S_n = \sum_{i=1}^n X_i$. Then $(V_i)_i = (X_i/S_n)_i \sim \text{Dir}(\alpha)$.

Proof. $S_n \sim \Gamma(\sum_{i=1}^n \alpha_i, \beta)$ by ?? and for $v \in \Delta^{n-1}$, we have

$$\begin{aligned} f_V(v) &= \int_0^\infty f_X(sv_1, \dots, sv_{n-1}, sv_n) f_{S_n}(s) ds \\ &= \int_0^\infty e^{-\sum_{i=1}^n sv_i} \left(\prod_{i=1}^n \frac{(sv_i)^{\alpha_i - 1}}{\Gamma(\alpha_i)} \right) \frac{s^{\sum_{i=1}^n \alpha_i - 1} e^{-s}}{\Gamma(\sum_{i=1}^n \alpha_i)} ds \\ &= \frac{1}{\prod_{i=1}^n \Gamma(\alpha_i)} \prod_{i=1}^n v_i^{\alpha_i - 1} \int_0^\infty e^{-s \sum_{i=1}^n v_i} s^{(\sum_{i=1}^n \alpha_i) - 1} ds \\ &= \frac{\Gamma(\sum_{i=1}^n \alpha_i)}{\prod_{i=1}^n \Gamma(\alpha_i)} \prod_{i=1}^n v_i^{\alpha_i - 1} \end{aligned}$$

□

Similar to ?? (??), where adding two Gammas yielded another Gamma where the parameters were added, Dirichlet distributions enjoy a similar kind of closure: “clumping” coordinate axes together (described below) yields another Dirichlet distribution where the parameters of the clumped axes are summed together.

Proposition 5 (Dirichlet clumping property)

Suppose $X \sim \text{Dir}(\alpha_1, \dots, \alpha_n)$. For any $r \leq n$, let $V_i = X_i$ for $i \in [r]$ and let $V_{r+1} = \sum_{j=r+1}^n X_j$. Then $V \sim \text{Dir}(\alpha_1, \dots, \alpha_r, \sum_{j=r+1}^n \alpha_j)$.

Proof. By induction, it suffices to show this for $r = n - 2$. Notice

$$\begin{aligned} f(v_1, \dots, v_r, s) &= B(\alpha)^{-1} \left(\prod_{i=1}^{n-1} v_i^{\alpha_i - 1} \right) \int \mathbb{1}\{x_{n-1} + x_n = s\} x_{n-1}^{\alpha_{n-1} - 1} x_n^{\alpha_n - 1} dx_{n-1} dx_n \\ &= B(\alpha)^{-1} \left(\prod_{i=1}^{n-1} v_i^{\alpha_i - 1} \right) \int_0^s u^{\alpha_{n-1} - 1} (s - u)^{\alpha_n - 1} du \\ &= B(\alpha)^{-1} \left(\prod_{i=1}^{n-1} v_i^{\alpha_i - 1} \right) s^{\alpha_{n-1} + \alpha_n - 1} B(\alpha_{n-1}, \alpha_n) \end{aligned}$$

Since $\frac{B(\alpha_{n-1}, \alpha_n)}{B(\alpha)} = \frac{\Gamma(\sum_{i=1}^n \alpha_i)}{\Gamma(\alpha_{n-1} + \alpha_n) \prod_{i=1}^{n-2} \Gamma(\alpha_i)}$, we are done. □

Iterating this result over coordinate axes other than the last $n - r$, we see that “clumping together” entries in a Dirichlet random vector yields another Dirichlet random vector with parameters also “clumped together.” Concretely, for any mapping $\phi : [n + 1] \rightarrow [m + 1]$ if $U_j = \sum_{\phi(i)=j} V_i$ then U has Dirichlet distribution with parameters $\gamma_j = \sum_{\phi(i)=j} \alpha_i$.

Generalizing this clumping property is the motivation for *Ferguson Distributions* [ferguson1973].

Definition 6 (Ferguson / Dirichlet process distribution)

Let μ be a finite positive Borel measure on complete separable metric space E . A random probability measure μ^* on E (i.e. a stochastic process indexed by a σ -algebra on E) has *Ferguson distribution with parameter μ* if for every finite partition $(B_i)_{i \in [r]}$ of E the random vector

$$(\mu^*(B_i))_{i \in [r]} \sim \text{Dir}(\mu(B_1), \dots, \mu(B_r))$$

Lemma 7 (Preservation of Ferguson under pushforward)

Let μ^* be Ferguson with parameter μ and $\phi : E \rightarrow F$ measurable. Then the pushforward $\mu^* \circ \phi^{-1}$ is a random probability measure on F that has Ferguson distribution with parameter $\mu \circ \phi^{-1}$.

Proof. For $(B_i)_{i \in [r]}$ a finite partition of F , $(\phi^{-1}(B_i))_i$ is a finite partition of E . Since μ^* is Ferguson

$$(\mu^*(\phi^{-1}(B_i)))_i \sim \text{Dir}((\mu(\phi^{-1}(B_i))))_i$$

Hence $\mu^* \circ \phi^{-1}$ is Ferguson with parameter $\mu \circ \phi^{-1}$. \square

Next, we turn to an important class of a Ferguson distributions arising from generalizing the Pólya urn.

Definition 8 (Polya sequence)

A sequence $(X_n)_{n \in \mathbb{N}}$ with values in E is a *Polya sequence with parameter μ* if for all $B \subset E$.

$$\begin{aligned} \Pr[X_1 \in B] &= \mu(B)/\mu(E) \\ \Pr[X_{n+1} \in B \mid X_1, \dots, X_n] &= \mu_n(B)/\mu_n(E) \end{aligned}$$

where $\mu_n = \mu + \sum_{i=1}^n \delta_{X_i}$.

Remark 9. When E is finite (e.g. a set of colors for the balls), (X_n) represents the result of successive draws from an urn with initially $\mu(x)$ balls of color $x \in E$ and after each draw a ball of the same color as the one drawn is added back to give an urn with color distribution $\mu_{n+1}(x)$.

[blackwell1973] gives the following result connecting Pólya sequences and Ferguson distributions.

Theorem 10 (Polya Urn Schemes)

Let (X_n) be a Polya sequence with parameter μ . Then:

1. $m_n = \mu_n/\mu_n(E)$ converges almost surely to a limiting discrete measure μ^*
2. μ^* has Ferguson distribution with parameter μ
3. Given μ^* , $(X_i)_{i \geq 1}$ are independent with distribution μ^*

Proof. First consider E finite. Let μ^* and $\{X_i\}$ be random variables whose joint distribution satisfies (2.) and (3.).

Let π_n be empirical distribution of $(X_i)_{i \in [n]}$. $X_i \stackrel{\text{iid}}{\sim} \mu^*$, so by SLLN $\pi_n \xrightarrow{a.s.} \mu^*$ and since

$$m_n = \frac{\mu + n\pi_n}{\mu(E) + n} \tag{2}$$

(1.) follows.

To complete the proof, we show equality in distribution of $\{X_i\}$ with a Polya- μ sequence. This amounts to showing

$$\Pr[A] = \prod_x \mu(x)^{n(x)} / \mu(E)^{[n]} \tag{3} \quad \{\text{eq:polya-seq-meas}\}$$

where $A = \{X_i = x_i\}_i \in \{0, 1\}^n$ and $n(x) = \#\{i : x_i = x\}$, and the rising factorial $a^{[k]} = a(a+1) \cdots (a+k-1)$. By the tower rule and $\{X_i\}$ IID

$$\Pr[A] = \mathbb{E}[\Pr[A \mid \mu^*]] = \mathbb{E}\left[\prod_x \mu^*(x)^{n(x)}\right] \tag{4}$$

Since μ^* is Ferguson, viewing $E = \sqcup_{x \in E} \{x\}$ as a partition we have $(\mu^*(x))_{x \in E} \sim \text{Dir}((\mu(x))_{x \in E})$ so the RHS is the $(n(x))_{x \in E}$ moment of the Dirichlet distribution, which is equal to

$$\mathbb{E} \left[\prod_x \mu^*(x)^{n(x)} \right] = \frac{\Gamma(\mu(E))}{\Gamma(\mu(E) + n)} \prod_x \frac{\Gamma(\mu(x) + n(x))}{\Gamma(\mu(x))} = \frac{1}{\mu(E)^{[n]}} \prod_x \mu(x)^{[n(x)]} \quad (5) \quad \{\text{eq:dirichlet-moment}\}$$

as required by ??.

General E follows from approximation argument. \square

Notice that the Dirichlet moment comparison in ?? was the key step relating μ to μ^* .

We leave the discreteness part of (1.) as an exercise, noting that similar to how Dirichlets can be defined as a set of independent Gammas normalized by their sum (?? (??)) we would expect the Dirichlet process / Ferguson random measures to be definable as a gamma process with independent “increments” divided by their sum.

Exercise 11. Prove every Ferguson random measure is discrete. (Hint: argue using moments).

Remark 12. If (X_i) is a Polya sequence, then it is a mixture of IID sequences (each drawn from μ^*) with mixture weights given by the Ferguson distribution on μ^* . Hence, (X_i) is exchangeable i.e. $(X_i) \stackrel{d}{=} (X_{\sigma(i)})$. This is already apparent in ??, and more generally de Finetti’s theorem guarantees that *any* exchangeable sequence is a mixture of IID sequences.

1.2 Construction of Haar Measure

For a finite group G , the measure $\mu(g) = \frac{1}{\#G}$ is left and right translation invariant i.e. $\mu(gA) = \mu(A) = \mu(Ag)$ for all $A \subset G$. As we will prove, all compact groups have unique translation invariant measure, called the Haar measure.

Example 13

Let $Z_{ij} \stackrel{\text{iid}}{\sim} N(0,1)$ for $i, j \in [n]$ and X the Gram-Schmidt orthonormalization of the rows of Z . By rotation invariance of Z , we can show $XU \stackrel{d}{=} UX$ for all $U \in O(n)$, so X has Haar measure on the compact (Lie) group $O(n)$.

Definition 14

A *topological vector space* (TVS) is a vector space equipped with a topology such that vector space operations are jointly continuous.

Example 15

\mathbb{R}^n with standard topology, any Banach space.

Definition 16

A family \mathfrak{G} of linear transformations on TVS \mathfrak{X} is (*uniformly*) *equicontinuous on subset* $K \subset \mathfrak{X}$ if for every neighborhood V of the origin, there exists a neighborhood U of the origin such that

$$\forall k_1, k_2 \in K : k_1 - k_2 \in U \Rightarrow \mathfrak{G}(k_1 - k_2) \subset V$$

That is, $T(k_1 - k_2) \in V$ for all $T \in \mathfrak{G}$.

We only need to verify at the origin because linearity of \mathfrak{G} and vector space structure allow us to translate the neighborhoods to any $p \in \mathfrak{X}$.

Remark 17. Whereas “uniform” is used in analysis to generalize the U neighborhood of continuity (e.g. the δ in ε - δ definition of continuity) from at a particular $x_0 \in \mathfrak{X}$ to $\forall x \in \mathfrak{X}$, “equi” is used to generalize from a single $f \in \mathcal{C}(\mathfrak{X})$ to a family $\mathfrak{G} \subset \mathcal{C}(\mathfrak{X})$.

Definition 18 (In-Class)

A *locally convex topological vector space* (LCTVS) is a TVS such that the topology has a base consisting of convex sets.

To construct Haar measure for any compact group, we will need a fix point theorem due to Kakutani.

Theorem 19 (Kakutani Fix Point Theorem)

K compact convex subset of LCTVS \mathfrak{X} , \mathfrak{G} group of linear transforms equicontinuous on K and such that $\mathfrak{G}(K) \subset K$, then $\exists p \in K$ such that

$$\mathfrak{G}(p) = \{p\} \quad (6)$$

Proof. Let P be the class of all non-empty compact convex subsets of K which are \mathfrak{G} -invariant, ordered by containment. $K \in P$ so P is not empty, and since any descending chain in P is lower bounded by the intersection of all of its elements (which is also in P) we may apply Zorn's lemma to conclude there is some minimal compact convex \mathfrak{G} -invariant $K_1 \subset K$. We are done if $K_1 = \{p\}$ is a singleton, so assume otherwise. We will contradict minimality of K_1 by constructing $K_2 \subsetneq K_1$ such that $K_2 \in P$.

We first exploit equicontinuity to construct a convex \mathfrak{G} -invariant open set U which approximates $K_1 - K_1$ (i.e. $(1 + \varepsilon)U \supset K_1 - K_1$ but $(1 - \varepsilon)\bar{U} \not\supset K_1 - K_1$):

- By assumption $K_1 - K_1$ (as a Minkowski sum) contains a point other than the origin, so because \mathfrak{X} Hausdorff there exists a neighborhood of the origin $V \in N(0)$ such that $\bar{V} \not\supset K_1 - K_1$.
- V may not be convex, but since \mathfrak{X} is a LCTVS there is convex V_1 in the local base of 0 such that $0 \in V_1 \subset V$.
- V_1 is convex, but not \mathfrak{G} -invariant. Note that \mathfrak{G} is a group so $\mathfrak{G}\mathfrak{G}A = \mathfrak{G}A$. To exploit this idea and construct a \mathfrak{G} -invariant convex open set, we first use equicontinuity of \mathfrak{G} on $K \supset K_1$ to obtain $U_1 \in N(0)$ such that $\mathfrak{G}(U_1 \cap (K_1 - K_1)) \subset \mathfrak{G}(U_1) \subset V_1$.
- Taking the convex hull (and exploiting convexity of V_1), we have

$$U_2 := \text{conv}(\mathfrak{G}(U_1 \cap (K_1 - K_1))) \subset \text{conv}(V_1) = V_1$$

U_2 is non-empty ($0 \in U_1 \cap (K_1 - K_1)$), relatively open in $K_1 - K_1$ ($T \in \mathfrak{G}$ invertible maps open sets to open sets), and $\mathfrak{G}U_2 = U_2$ because:

- T is linear so $T \text{conv}(A) = \text{conv}(TA)$
- $T \in \mathfrak{G}$ invertible (\mathfrak{G} is a group) so $T(A \cap B) = TA \cap TB$ for sets A, B .
- $\mathfrak{G}\mathfrak{G}A = \mathfrak{G}A$.

By continuity, $\mathfrak{G}U_2 = \overline{\mathfrak{G}U_2}$.

- Let $U := \delta U_2$ where $\delta = \inf\{a : a > 0, aU_2 \supset K_1 - K_1\}$, by compactness of $K_1 - K_1$ we have $\delta < \infty$. With this definition, for any $\varepsilon \in (0, 1)$

$$(1 + \varepsilon)U \supset K_1 - K_1 \not\subset (1 - \varepsilon)\bar{U}$$

Note that equicontinuity was required to bound $U_2 \subset V_1$.

Next, we will exploit $(1 - \varepsilon)\bar{U} \not\supset K_1 - K_1$ to construct a proper subset $K_2 \subsetneq K_1$ and use relative openness of U and compactness of K_1 to argue non-emptiness by constructing $p \in K_2$, contradicting minimality.

- For the relatively open cover $\{2^{-1}U + k\}_{k \in K_1}$ of K_1 , let $\{k_i\}_{i=1}^n$ index a finite subcover and define (the center) $p = \frac{1}{n} \sum_{i=1}^n k_i$. Notice $p \in K_1$ (by convexity of K_1) and every $k \in K_1$ satisfies $k \in k_i + 2^{-1}U$ for some $i \in [n]$. Additionally, $k \in k_j + (1 + \varepsilon)U$ for all j (since $K_1 - K_1 \subset (1 + \varepsilon)U$) so we have

$$p \in \frac{1}{n} (2^{-1}U + (n - 1)(1 + \varepsilon)U) + k$$

Setting $\varepsilon = \frac{1}{4(n-1)}$ we get $p \in (1 - \frac{1}{4n})U + k$ for each $k \in K_1$, i.e. every point in K_1 is within $(1 - \frac{1}{4n})U$ of the “center” p .

- As p is within a $(1 - 1/4n)\bar{U}$ ball of every $k \in K_1$ we can define the non-empty set

$$K_2 = K_1 \cap \bigcap_{k \in K_1} \left(\left(1 - \frac{1}{4n}\right)\bar{U} + k \right) \supset \{p\} \neq \emptyset$$

Because $(1 - \frac{1}{4n})\bar{U} \not\supset K_1 - K_1$, we have $K_2 \subsetneq K_1$ is a proper subset.

- To contradict minimality of K_1 , it remains to verify K_2 satisfies the desired properties. K_2 is closed and convex because (due to how we constructed \bar{U}) it is the intersection of closed convex sets. Further, since $T(a\bar{U}) \subset a\bar{U}$ for $T \in \mathfrak{G}$, we have

$$T(a\bar{U} + k) \subset a\bar{U} + Tk \quad \text{for all } T \in \mathfrak{G}, k \in K_1$$

Combined with $\mathfrak{G}K_1 \subset K_1$ and $Tk \in K_1$ for all $k \in K_1$, we have that $\mathfrak{G}K_2 \subset K_2$. □

2 Lecture 2: Measure theory

2020-01-23

Throughout our discussion, all topological spaces are assumed Hausdorff unless explicitly noted otherwise.

2.1 Construction of Haar measure

Definition 20

A *topological group* is a group equipped with a topology such that the group operations $(g, h) \mapsto gh$ and $g \mapsto g^{-1}$ are continuous.

Theorem 21 (*Existence of Haar Measure*)

Let G be a compact topological group and $\mathcal{C}(G)$ the set of continuous maps $G \rightarrow \mathbb{R}$. Then there is a unique linear form $m : \mathcal{C}(G) \rightarrow \mathbb{R}$ such that

1. $m(f) \geq 0$ for $f \geq 0$ (positive)
2. $m(1) = 1$ (normalized)
3. $m({}_s f) = m(f)$ where ${}_s f(g) = f(s^{-1}g)$ for $s, g \in G$ (left invariant)
4. $m(f_s) = m(f)$ where $f_s(g) = f(gs)$ (right invariant)

m is called the Haar measure on G .

We will need the following theorem to relate compactness with equicontinuity:

Theorem 22 (*Generalization of Arzela-Ascoli*)

Let X be a compact Hausdorff space. A subset of \mathbb{R} -valued continuous functions $F \subset \mathcal{C}(X)$ is relatively compact in topology induced by uniform norm $\|\cdot\|_\infty \Leftrightarrow F$ is equicontinuous and pointwise bounded.

Proof of ??. Fix $f \in \mathcal{C}(G)$ and let \mathcal{C}_f denote the convex hull of all left translates of f , i.e. $g \in \mathcal{C}_f$ are finite sums of form

$$g(x) = \sum_{\text{finite}} a_i f(s_i x), \quad a_i > 0, \sum_{\text{finite}} a_i = 1, s_i \in G$$

Clearly $\|g\|_\infty \leq \|f\|_\infty < \infty$, thus $\mathcal{C}_f(x) = \{g(x) : g \in \mathcal{C}_f\}$ is bounded for all $x \in G$ hence \mathcal{C}_f is pointwise bounded.

As f is a continuous function on compact G , it is uniformly continuous hence for $\varepsilon > 0$ there exists a neighborhood V_ε of the identity $e \in G$ such that

$$y^{-1}x \in V_\varepsilon \Rightarrow |f(x) - f(y)| \leq \varepsilon$$

Since $(s^{-1}y)^{-1}s^{-1}x = y^{-1}x$, we also have

$$y^{-1}x \in V_\varepsilon \Rightarrow |{}_s f(y) - {}_s f(x)| < \varepsilon$$

Since $g \in \mathcal{C}_f$ are convex combinations of ${}_s f$, by the triangle inequality

$$y^{-1}x \in V_\varepsilon \Rightarrow |g(y) - g(x)| < \varepsilon$$

As this works for any $g \in \mathcal{C}_f$, we have that \mathcal{C}_f is equicontinuous.

By ?? (??), \mathcal{C}_f is relatively compact in $\mathcal{C}(G)$, so its closure $K_f := \overline{\mathcal{C}_f}$ is compact (and still convex).

Consider G acting on $\mathcal{C}(G)$ by left translation $f \mapsto {}_s f$. Notice $G\mathcal{C}_f \subset \mathcal{C}_f$ (as \mathcal{C}_f already contains all finite convex combinations of all left translations of f) and hence $GK_f \subset K_f$ as well.

Furthermore, $\|_s f - {}_s g\|_\infty = \|f - g\|_\infty$ so G acts as a group of isometries on $\mathcal{C}(G)$. In particular, this group is equicontinuous (with the same $U = V$ in ??).

Taking $\mathfrak{G} = G$ and $K = K_f$ in ?? (??), there is a fixed point $g \in K_f$ of this action of G on K_f which satisfies

$${}_s g = g \quad (\forall s \in G) \quad \Rightarrow \quad g(s^{-1}) = {}_s g(e) = g(e) = c \quad (\forall s \in G)$$

for some constant $c \in \mathbb{R}$ (which we will later use to define $m(f) := c$).

We first show there is only one constant function in K_f , so the fix point $Gg = \{g\} = \{c\mathbb{1}\}$ is unique and $m(f) = c$ is well defined. For any constant function $c\mathbb{1} \in K_f$ and $\varepsilon > 0$, we can (because $K_f = \overline{\mathcal{C}_f}$) find $\{s_1, \dots, s_n\} \subset G$ and $a_i > 0$ such that

$$\sum_{i=1}^n a_i = 1, \quad \text{and} \quad \left| c - \sum_{i=1}^n a_i f(s_i x) \right| < \varepsilon \quad (\forall x \in G) \quad (7) \quad \{\text{eq:combo-close-to-constant}\}$$

for any $\varepsilon > 0$.

Similarly, consider the same construction as before expect now use right translations of f (i.e. using the opposite group G' of G , or the function $f' = f(x^{-1})$, obtaining relatively compact set \mathcal{C}'_f with compact convex closure K'_f with fix point $g' = c'\mathbb{1}$). Approximating $c'\mathbb{1}$ using \mathcal{C}'_f , we have

$$\left| c' - \sum_j b_j f(x t_j) \right| < \varepsilon \quad (\text{for some } t_j \in G, b_j > 0 \text{ with } \sum_j b_j = 1) \quad (8)$$

Opposite group

The opposite group g' of the group G is the group that coincides with G as a set but has group operation $(x, y) \mapsto y^{-1}x^{-1}$

Summing over i

$$\left| c' - \sum_{i,j} a_i b_j f(s_i t_j) \right| < \varepsilon \sum_i a_i = \varepsilon$$

Operating symmetrically on ?? (multiply by b_i and put $x = t_i$) shows

$$\left| c - \sum_{i,j} a_i b_j f(s_i t_j) \right| < \varepsilon$$

Together, we have $|c' - c| < 2\varepsilon$ so taking $\varepsilon \rightarrow 0$ shows $c' = c$. Since $c\mathbb{1} \in K_f$ was an arbitrary constant function, we have that the constant function in K_f is actually unique and so the function $m(f) := c \in K_f$ is well defined. Moreover, $m(f)\mathbb{1}$ is the *only* constant function which can be arbitrary well approximated by convex combinations of left or right translates of f .

The following properties are obvious:

- $m(\mathbb{1}) = 1$ since $K_f = \{1\}$ for $f = \mathbb{1}$
- $m(f) \geq 0$ if $f \geq 0$
- $m({}_s f) = m(f) = m(f_s)$ (since $K_{{}_s f} = K_f$, $K'_f = K'_{f_s}$, and uniqueness of $m(f)\mathbb{1}$ being the only constant function approximable by both K_f and K'_f)
- $m(af) = am(f)$ for any $a \in \mathbb{R}$ (since $K_{af} = K_f$)

To show m is linear, it suffices (due to the last bullet above) to show that m is additive. Fix $f, g \in \mathcal{C}(G)$. Approximate $m(f)$ using K_f to get

$$\left| m(f) - \sum_{\text{finite}} a_i f(s_i x) \right| \quad (9) \quad \{\text{eq:approx-mf-using-Kf}\}$$

Define $h(x) = \sum_{\text{finite}} a_i g(s_i x)$ using the same a_i and s_i and approximate $m(h)$ using \mathcal{C}_h to get

$$\left| m(h) - \sum_{\text{finite}} b_j h(t_j x) \right| < \varepsilon$$

Since $h \in \mathcal{C}_g$, we have $\mathcal{C}_h \subset \mathcal{C}_g$ hence $K_h \subset K_g$. But $m(g)\mathbb{1} \in K_g$ is the only constant function so $m(h) = m(g)$ and (after expanding the definition of h) we have

$$\left| m(g) - \sum_{i,j < \infty} a_i b_j g(s_i t_j x) \right| < \varepsilon$$

On the other hand, multiplying ?? by b_j replacing x with $t_j x$, summing over j , and finally adding with the above inequality gives

$$|m(f) + m(g) - \sum_{i,j} a_i b_j (f + g)(s_i t_j x)| < 2\varepsilon$$

Thus $m(f) + m(g) \in K_{f+g}$, establishing additivity. Note that the only constant in K_{f+g} is $(m(f) + m(g))\mathbb{1}$. \square

2.2 Facts from topology

We now want to head towards some integration against probability measures defined on spaces more abstract than \mathbb{R}^n .

Definition 23

A topological space X is *normal* if for any disjoint closed sets Y and Z there exists disjoint open sets U and V such that $Y \subset U$ and $Z \subset V$.

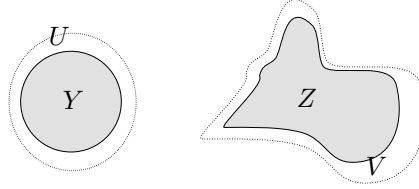


Figure 1: Normal topological spaces admit separating closed sets with two disjoint open sets

Definition 24

X is *completely regular* (*Tychonoff* if X is also Hausdorff) if for all $y \in X$ and every closed $Z \subset X \setminus \{y\}$ there exists $f : X \rightarrow [0, 1]$ continuous such that $f(y) = 0$ and $f(z) = 1$ for all $z \in Z$. We say y and Z are separated by a (Urysohn) function.

Corollary 25 (Urysohn's Lemma)

Every normal space is completely regular.

Lemma 26

A compact (Hausdorff) space is normal hence completely regular.

Proof. Fix disjoint closed Y and Z and let $y \in Y$. Consider the open cover of Z given by $\{V_{y,z} : z \in Z\}$ where each $V_{y,z} \in N(z)$ is disjoint from some $U_{y,z} \in N(y)$ (existence ensured by Hausdorff). By compactness, there exists a finite subcover $\{V_{y,z_i}\}_{i=1}^n$. For each of these V_{y,z_i} , let $U_{y,z_i} \in N(y)$ denote the corresponding disjoint neighborhood of y and consider

$$U'_y = \bigcap_{i=1}^n U_{y,z_i} \in N(y)$$

U'_y is open because it is the intersection of finitely many open sets. It is also disjoint from

$$V'_y := \bigcup_{i=1}^n V_{y,z_i}$$

which contains B and is also open.

Now consider the open cover $\{U'_y : y \in Y\}$, let $\{U'_{y_i}\}_{i=1}^n$ be a finite subcover, and let $U = \bigcup_{i=1}^n U'_{y_i}$. Analogously, let $V = \bigcap_{i=1}^n V'_{y_i}$ where V'_y is given above (open cover of B and disjoint from U'_y). Then $U \supset Y$ and $V \supset Z$ provide two disjoint separating open sets. \square

Lemma 27

A topological space (X, τ) is completely regular (i.e. Tychonoff) space iff the original topology coincides with the initial topology $\tau(X, \mathcal{C}(X))$ i.e. the smallest topology that makes every function in $\mathcal{C}(X)$ continuous.

Proof. We only show \Rightarrow . Let U be τ -open and for $x \in U$ pick an Urysohn function $f \in \mathcal{C}(X)$ such that $f(x) = 0$ and $f(U^c) = 1$. Then $V_x = \{y : f(y) < 1\} = f^{-1}((-\infty, 1))$ is a $\sigma(X, \mathcal{C}(X))$ -open neighborhood of x contained in U , so $U = \bigcup_{x \in U} V_x$ is $\sigma(X, \mathcal{C}(X))$ -open. Since $\sigma(X, \mathcal{C}(X))$ is minimal, we have $\tau = \sigma(X, \mathcal{C}(X))$. \square

2.3 Radon, Borel, and Baire measures

Definition 28

A non-negative set function $m : 2^X \rightarrow [0, +\infty]$ on X is an *outer measure on X* (or Carathéodory outer measure) if:

1. $m(\emptyset) = 0$
2. $A \subset B \Rightarrow m(A) \leq m(B)$ (monotone)
3. $m(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} m(A_n)$ for all $A_n \subset X$. (countable subadditivity)

Outer measures are over-approximations of the “size” of A . See ??, where we see that $m(A)$ is obtained using an over-approximation $A \subset \bigcup_{n=1}^{\infty} X_n \in \mathfrak{X}$.

Definition 29

Let $m : 2^X \rightarrow [0, +\infty]$ be a non-negative set function satisfying $m(\emptyset) = 0$. A set $A \subset X$ is *Carathéodory measurable wrt m* (Carathéodory m -measurable) if for any $E \subset X$

$$m(E) = m(E \cap A) + m(E \setminus A)$$

We use \mathfrak{M}_m to denote the class of all Carathéodory m -measurable sets.

It turns out m enjoys nice properties when restricted to \mathfrak{M}_m , and when m is an outer measure we end up with a countably additive function defined on a σ -algebra! The below theorem is one way of arriving at the Lebesgue measure (although we will be using to define Daniell integration).

Theorem 30 (Carathéodory construction)

1. \mathfrak{M}_m is an algebra, m is additive on \mathfrak{M}_m
2. (Finite additivity) For all sequences of pairwise disjoint $A_i \in \mathfrak{M}_m$ and any $E \subset X$

$$\begin{aligned} m\left(E \cap \bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n m(E \cap A_i) \\ m\left(E \cap \bigcup_{i=1}^{\infty} A_i\right) &= \sum_{i=1}^{\infty} m(E \cap A_i) + \lim_{n \rightarrow \infty} m\left(E \cap \bigcup_{i=n}^{\infty} A_i\right) \end{aligned}$$

3. If m is an outer measure on X , then \mathfrak{M}_m is a σ -algebra, m is countably additive on \mathfrak{M}_m , and m is complete (subsets of null sets also have measure zero) on \mathfrak{M}_m

Remark 31. The outer measure is constructed such that it satisfies countable additivity on the measurable sets \mathfrak{M}_m .

Example 32 (Munroe construction of outer measure)

Let \mathfrak{X} be a family of subsets of X such that $\emptyset \in \mathfrak{X}$. Given $\tau : \mathfrak{X} \rightarrow [0, +\infty]$ with $\tau(\emptyset) = 0$, set

$$m(A) = \inf \left\{ \sum_{n=1}^{\infty} \tau(X_n) : X_n \in \mathfrak{X}, A \subset \bigcup_{n=1}^{\infty} X_n \right\}$$

where $m(A) = \infty$ in the absence of such sets X_n . Then m is an outer measure, denoted τ^* .

This is where the “outer” comes from: $\bigcup_n X_n \supset A$ is an outer approximation to A using (potentially overlapping) sets from \mathfrak{X} hence $\sum_{n=1}^{\infty} \tau(X_n)$ is an overapproximation to the “size” of A . $m(A)$ is the best (i.e. smallest) overapproximation.

Recall the *Borel σ -algebra*, denoted $\mathcal{B}(X)$, is generated by all open sets.

Definition 33 (Baire σ -algebra)

A *functionally open set* is of the form

$$\{x \in X : f(x) > 0\}, \quad \text{for } f \in \mathcal{C}(X) \quad (10)$$

The *Baire σ -algebra*, denoted by $\mathcal{Ba}(X)$, is the σ -algebra generated by functionally open sets. Elements of $\mathcal{Ba}(X)$ are called *Baire sets* in X .

Remark 34. $\mathcal{Ba}(X)$ is the smallest σ -algebra where every $f \in \mathcal{C}(X)$ is measurable. It coincides (via a truncation and monotonicity argument) to the smallest one making every $f \in \mathcal{C}_b(X)$ measurable. Contrast this to ??, which shows that completely regular spaces are those with the smallest topology where every $f \in \mathcal{C}(X)$ is continuous.

Remark 35. Since the functionally open sets can be written as $f^{-1}((0, \infty))$ for continuous f , they are also Borel sets. Therefore, the class of Baire sets are contained in the class of Borel sets.

Lemma 36

In a metric space (X, d) , any closed set S is the set of zeros of a continuous function (namely $d_S(x) = \inf_{s \in S} d(x, s)$). Hence, $\mathcal{B}(X) = \mathcal{Ba}(X)$.

Lemma 37 (Baire sets are countably determined)

Every $A \in \mathcal{Ba}(X)$ is determined by some countable family of functions, i.e. has the form

$$A = \{x : (f_i(x))_{i=1}^{\infty} \in B\} \quad \text{for some } f_i \in \mathcal{C}(X), B \in \mathcal{B}(\mathbb{R}^{\mathbb{N}_0})$$

Moreover, every set of this form is Baire and we can take $f_i \in \mathcal{C}_b(X)$.

Proof. We first show every set of the same form as A is Baire. True if B is closed, since ?? allows us to write $B = \phi^{-1}(0)$ for some continuous function $\phi : \mathbb{R}^{\mathbb{N}_0} \rightarrow \mathbb{R}$ so $\psi = x \mapsto \phi((f_n(x))_{n \geq 1})$ is continuous hence $A = \psi^{-1}(0)$ is also closed. For any fixed $\{f_n\}_{n \geq 1}$, the class of sets $B \in \mathcal{B}(\mathbb{R}^{\mathbb{N}_0})$ satisfying

$$\{x : (f_i(x))_{i \geq 1} \in B\} \in \mathcal{Ba}(X)$$

is a σ -algebra containing $B = \prod_i (-\infty, a_i)$ where $a_i \neq \infty$ for only finitely many i . This is a basis for $\mathcal{B}(\mathbb{R}^{\mathbb{N}_0})$, thus $\mathcal{Ba}(X)$ contains it and the two coincide (recall $\mathcal{Ba} \subset \mathcal{B}$ since functionally determined sets are \mathcal{B} -open).

On the other hand, the class \mathcal{E} of all Baire sets E representable like A with $f_i \in \mathcal{C}_b(X)$ contains the functionally open sets. It is also a σ -algebra, since for $E \in \mathcal{E}$:

- We can represent E^c using the same $\{f_i\}$ and B_j^c instead.
- $E = \bigcap_{j=1}^{\infty} E_j$ can be represented by embedding all the $\{f_i\}$ and $\{B_j\}$ for each of the countably many E_j into a single countably infinite sequence (i.e. $B = \prod_{j=1}^{\infty} B_j$).

□

The following is a useful consequence of Dynkin's π - λ theorem applied to simplify determining when two Borel measures are equal. More generally, π - λ allows us to verify a property on a class of sets \mathcal{E} closed under finite intersection (π -system) and conclude the property on the more complicated σ -algebra $\sigma(\mathcal{E})$ provided that the set

$$D = \{A \in \sigma(\mathcal{E}) : A \text{ satisfies the property}\}$$

is a λ -system (closed under complement and disjoint unions).

Lemma 38

If two probability measures agree on a class of sets \mathcal{E} closed under finite intersections, then they also coincide on the σ -algebra generated by \mathcal{E} .

Proof. By hypothesis \mathcal{E} is a π -system and the class $D = \{A : \mu(A) = \nu(A)\}$ is a λ -system (by properties of a probability measure) so the result follows from Dynkin's π - λ theorem. □

Throughout, we consider (signed) measures of *bounded variation* unless explicitly denoted otherwise. This means that

$$|\mu|(\Omega) = \mu^+(\Omega) + \mu^-(\Omega) < \infty$$

Definition 39

Let X be a topological space.

- A countably additive measure on $\mathcal{B}(X)$ is called a *Borel measure*
- A countably additive measure on $\mathcal{B}a(X)$ is called a *Baire measure*
- A Borel measure μ on X is called a *Radon measure* if every $B \in \mathcal{B}(X)$ can be approximated from the inside by compact sets: for $\varepsilon > 0$ exists $K_\varepsilon \subset B$ such that $|\mu|(B \setminus K_\varepsilon) < \varepsilon$.

Lemma 40

If two Borel measures coincide on all open sets, then they coincide on all Borel sets.

Proof. By taking differences, it suffices to verify μ vanishing on open sets must be identically zero. Split $\mu = \mu^+ - \mu^-$ and notice that each of the two components are nonnegative and coincide on open sets. As open sets are closed under finite section and \mathcal{B} is generated by open sets, the results follows from ?? . □

We now move from Borel measures to Radon measures. First observe by definition that μ is Radon iff $|\mu|$ is Radon iff both μ^+ and μ^- are Radon, so we only really need to study when non-negative Borel measures $\mu \geq 0$ are Radon. As the study of Radon measures will inevitably require inner approximation by compact sets, we first consider the case of $X = \mathbb{R}^n$.

Theorem 41 (Open/compact approximation on metric spaces)

Let $\mu \geq 0$ be a Borel measure on a metric space. Then for any Borel set B and $\varepsilon > 0$, there exists U_ε open and K_ε compact such that $K_\varepsilon \subset B \subset U_\varepsilon$ and $\mu(U_\varepsilon \setminus K_\varepsilon) < \varepsilon$.

Thus, Borel measures are Radon on metric spaces.

Proof. Fix $\varepsilon > 0$. It suffices to show there exists closed $F_\varepsilon \subset B$ such that $\mu(B \setminus F_\varepsilon) < \varepsilon/2$, since then $K_\varepsilon = F_\varepsilon \cap \bar{B}_r(0)$ (r sufficiently large, exists because μ bounded variation) is a compact set approximating F_ε within $\varepsilon/2$ and additivity of μ completes the proof.

Let \mathcal{A} denote the class of all sets $A \in \mathcal{B}$ such that $F_\varepsilon \subset A \subset U_\varepsilon$ and $\mu(U_\varepsilon \setminus F_\varepsilon) < \varepsilon$ for some closed set F_ε and open set U_ε . Every closed A is in \mathcal{A} , since we can take $F_\varepsilon = A$ and $U_\varepsilon = \cup_{p \in U} B_\delta(p)$. with δ sufficiently small. Since the closed sets generate \mathcal{B} , it suffices to show \mathcal{A} is a σ -algebra. As \mathcal{A} is closed wrt complements (swap $U_\varepsilon = F_\varepsilon^c$ and vice versa), it remains to verify closure under countable union.

Fix $\varepsilon > 0$. For $j \in \mathbb{N}$ and $A_j \in \mathcal{A}$, there exists closed F_j and open U_j such that $F_j \subset A_j \subset U_j$ and $\mu(U_j \setminus F_j) < \varepsilon 2^{-j}$. $U = \cup_{j=1}^\infty U_j$ is open, $Z_k = \cup_{j=1}^k F_j$ is closed for all $k \in \mathbb{N}$, $Z_k \subset \cup_{j=1}^\infty A_j \subset U$, and

$$\mu(\cup_{j=1}^\infty (U_j \setminus F_j)) < \sum_{j=1}^\infty \varepsilon 2^{-j} = \varepsilon$$

While Z_k is only closed for finite k , notice $\mu(Z_k) \rightarrow \mu(\cup_{j=1}^{\infty} F_j)$ so by countable additivity

$$\varepsilon > \mu(\cup_{j=1}^{\infty} (U_j \setminus F_j)) = \mu(\cup_{j=1}^{\infty} U_j) - \mu(\cup_{j=1}^{\infty} F_j) \geq \mu(\cup_{j=1}^{\infty} U_j) - \mu(Z_k) - \varepsilon/2 = \mu(\cup_{j=1}^{\infty} U_j \setminus Z_k) - \varepsilon/2$$

for sufficiently large k . □

Definition 42

A nonnegative set function μ defined on a class \mathcal{A} of subsets of a topological space X is *tight* on \mathcal{A} if $\forall \varepsilon > 0$ exists compact $K_\varepsilon \subset X$ such that $\mu(A) < \varepsilon$ for all $A \in \mathcal{A}$ that does not meet K_ε .

An additive set function μ of bounded variation on an algebra is *tight* if its total variation $|\mu|$ is tight.

Tightness is important because it says the whole space is inner approximable by a compact set. Indeed, a Borel measure is tight iff $\forall \varepsilon > 0$ exists compact K_ε such that $|\mu|(X \setminus K_\varepsilon) < \varepsilon$.

The above definition is more general because in the case of a general Baire measure, nonempty compact sets may not belong to the domain of μ so $\mu(X \setminus K_\varepsilon) = \mu(X) - \mu(K_\varepsilon)$ may not be measurable.

If instead of inner approximation by compact K_ε , we consider approximation by closed sets F_ε and insist the error $A \setminus F_\varepsilon$ remains measurable, then we arrive at the following definition:

Definition 43

A nonnegative set function μ defined on a class \mathcal{A} of subsets of a topological space X is *regular* if $\forall A \in \mathcal{A}$ and $\varepsilon > 0$, $\exists F_\varepsilon$ closed such that $F_\varepsilon \subset A$, $A \setminus F_\varepsilon \in \mathcal{A}$, and $\mu(A \setminus F_\varepsilon) < \varepsilon$.

From ?? (??), we have that any Borel measures on metric spaces can be inner approximated by compacts so (after intersecting with $\bar{B}_r(0)$ for sufficiently large r) Borel measures of metric spaces are regular.

Corollary 44 (Baire measures are regular)

Every Baire measure μ on a topological space X is regular. Moreover, for every Baire set E and $\varepsilon > 0$, there exists a continuous function f on X such that $f^{-1}(0) \subset E$ and $|\mu|(E \setminus f^{-1}(0)) < \varepsilon$.

Proof. Idea: exploit ?? (??) to move focus to pushforward measure on metric space \mathbb{R}^∞ .

By splitting $\mu = \mu^+ - \mu^-$, it suffices to consider non-negative measures. By ??, E is of the form

$$E = \{x : (f_i(x))_{i \geq 1} \in B\}$$

where $f_i \in \mathcal{C}(X)$ and $B \in \mathcal{B}(\mathbb{R}^\infty)$. Define the continuous function $g(x) = (f_i(x))_{i \geq 1}$ from X to \mathbb{R}^∞ and consider the pushforward measure $g_*(\mu)$. It is a Borel measure on a metric space, so by ?? there exists closed $H \subset B$ such that $g_*(\mu)(B \setminus H) \leq \varepsilon$. Moreover, by ?? (??) there is some $h \in \mathcal{C}(\mathbb{R}^\infty)$ such that $H = h^{-1}(0)$. Finally, notice $f = h \circ g \in \mathcal{C}(X)$ and

$$\varepsilon > g_*(\mu(B \setminus h^{-1}(0))) = \mu(g^{-1}(B)) - \mu((g^{-1} \circ h^{-1})(0)) = \mu(E) - \mu(f^{-1}(0)) = \mu(E \setminus f^{-1}(0))$$

□

3 Lecture 3: Daniell integration

2020-01-28

Theorem 45 (Extension to Radon measure)

Suppose an algebra \mathcal{A} of subsets of topological space X contains a base of the topology. Let μ be a regular additive set function of bounded variation on \mathcal{A} . If μ is tight, then it admits a unique extension to a Radon measure on X .

Proof. V.I. Bogachev, “Measure Theory” Theorem 7.3.2 □

Corollary 46 (Tight Baire measures extend to Radon)

Let X be a completely regular space. Then every tight Baire measure μ on X admits a unique extension to a Radon measure.

Proof. Every Baire measure is regular by ??.

Since X is completely regular, by ?? its topology coincides with $\tau(X, \mathcal{C}(X))$: the smallest making every function in $\mathcal{C}(X)$ continuous. The functionally open sets form a base of this topology (they are the pullback of the base of open intervals for \mathcal{B} under all continuous functions $\mathcal{C}(X)$), so ?? yields the desired extension. \square

Definition 47

A *vector lattice of functions* is a linear space of real functions on a nonempty set Ω such that $\max(f, g) \in \mathcal{F}$ for all $f, g \in \mathcal{F}$.

Remark 48. Notice $\min(f, g) = \max(-f, -g) \in \mathcal{F}$ and $|f| \in \mathcal{F}$. Also, since $\max(f, g) = (|f - g| + f + g)/2$ it suffices to require \mathcal{F} be closed under absolute values.

Theorem 49 (Daniell integration)

Let \mathcal{F} be a vector lattice of functions on a set Ω such that $\mathbb{1} \in \mathcal{F}$. Let L be a linear functional on \mathcal{F} with:

- $L(f) \geq 0$ for all $f \geq 0$ (positive)
- $L(\mathbb{1}) = 1$
- $L(f_n) \rightarrow 0$ for every $f_n \downarrow 0$

Then there exists a unique probability measure μ on $\mathcal{A} = \sigma(\mathcal{F})$ generated by \mathcal{F} such that $\mathcal{F} \subset \mathcal{L}^1(\mu)$ and

$$L(f) = \int_{\Omega} f d\mu, \quad \forall f \in \mathcal{F}$$

Compare this with Riesz representation theorem

For X a compact space, L linear functional on $\mathcal{C}(X)$ with $L(\mathbb{1}) = 1$ and $L(f) \geq 0$ for $f \geq 0$ (positive linear functional), then $L(f) = \int_X f d\mu$ with unique regular Borel probability measure μ on X .

The relation is through Dini's theorem: If $\{f_n\} \subset \mathcal{C}(X)$, X compact, and $f_n(x) \downarrow 0$, then $\lim_{n \rightarrow \infty} \sup_{x \in X} f_n(x) = 0$.

Proof. Denote \mathcal{L}^+ the set of all bounded functions f of the form $f(x) = \lim_{n \rightarrow \infty} f_n(x)$, where $f_n \in \mathcal{F}$ are nonnegative and the sequence $\{f_n\}$ is increasing. $\{f_n\}$ is uniformly bounded, hence $\{L(f_n)\}$ is increasing and bounded by properties of L so by monotone convergence $\lim_n L(f_n)(x)$ exists for all x and we can extend L to $f \in \mathcal{L}^+$ by defining $L(f) = \lim_n L(f_n)$.

We show that the extended functional $L(f)$ is well-defined, coincides on bounded nonnegative functions in \mathcal{F} with the original functional, and possesses the following properties:

1. $L(f) \leq L(g)$ for all $f, g \in \mathcal{L}^+$ with $f \leq g$ (positive)
2. $L(f + g) = L(f) + L(g)$, $L(cf) = cL(f)$ for all $f, g \in \mathcal{L}^+$ and $c \in [0, +\infty)$ (linear)
3. $\min(f, g) \in \mathcal{L}^+$, $\max(f, g) \in \mathcal{L}^+$, and

$$L(f) + L(g) = L(\min(f, g)) + L(\max(f, g))$$

for all $f, g \in \mathcal{L}^+$

4. $\lim_n f_n \in \mathcal{L}^+$ for every uniformly bounded increasing sequence of functions $f_n \in \mathcal{L}^+$, and $L(\lim_n f_n) = \lim_n L(f_n)$.

Suppose $\{f_n\}$ and $\{g_k\}$ are two increasing sequences of nonnegative functions in \mathcal{F} with $\lim_n f_n \leq \lim_k g_k$. Then $\min(f_n, g_k) \in \mathcal{F}$ are increasing to f_n (because $f_n \leq \lim_n f_n \leq \lim_k g_k$) hence

$$L(f_n) = \lim_k L(\min(f_n, g_k)) \leq \lim_k L(g_k)$$

where the first equality follows from properties of L (take difference between successive $k+1$ and k terms, use linearity and positivity and decreasing residual term) and the second because $g_k - \min(f_n, g_k) \geq 0$ for all k and L is positive and linear. Take $n \rightarrow \infty$ to conclude $\lim_n L(f_n) \leq \lim_k L(g_k)$.

If $\{f_n\}$ and $\{g_k\}$ both converge to the same $f \in \mathcal{L}^+$, then apply the above result symmetrically to get $\lim_n L(f_n) = \lim_k L(g_k)$, hence L is well-defined on \mathcal{L}^+ . By considering constant sequences for $f \in \mathcal{F}$, we have that L coincides with the original on $\mathcal{F} \cap \mathcal{L}^+$ and therefore properties (1) and (2) continue to hold by linearity.

Property (3) is because for $f_n \uparrow f$ and $g_n \uparrow g$, $\mathcal{F} \ni \min(f_n, g_n) \uparrow \min(f, g) \in \mathcal{L}^+$ (analogously for max) and property (2) applied to

$$f + g = \min(f, g) + \max(f, g)$$

To verify (4), suppose $\mathcal{F} \ni f_{m,n} \uparrow f_m \in \mathcal{L}^+$ (note the sequence $\{f_m\}$ is not in \mathcal{F} , but each term is a limit of a sequence $\{f_{m,n}\}_n$ in \mathcal{F}). Let $g_m = \max_{n \leq m} f_{m,n} \in \mathcal{F}$, so $g_m \leq g_{m+1}$ and $f_{m,n} \leq g_m \leq f_m$ for $n \leq m$. Taking $n, m \rightarrow \infty$ shows $\lim_m f_m = \lim_m g_m \in \mathcal{L}^+$ so by well-definedness

$$\lim_m L(f_m) = \lim_m L(g_m) = L(\lim_m g_m)$$

But since g_m and f_k are both increasing, $\lim_k f_k - g_m \downarrow 0$ so in fact (by property of L)

$$\lim_m L(f_m) = L(\lim_m g_m) = L(\lim_k f_k)$$

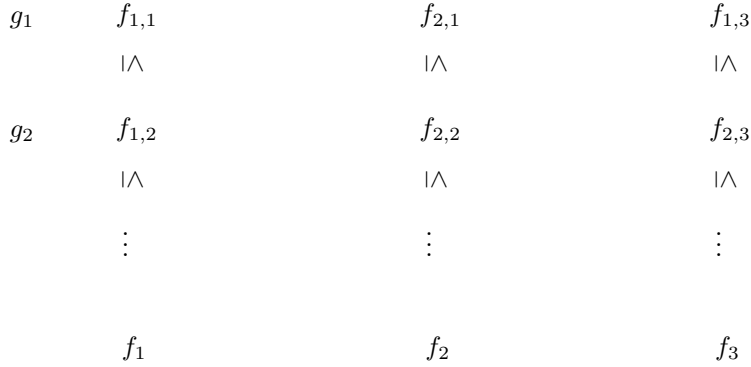


Figure 2: Sketch of the inequalities involved in proving property (4)

Armed with this extension of L to \mathcal{L}^+ , we now define μ . Denote by \mathcal{G} the class of all sets $G \subset \Omega$ with $\mathbb{1}_G \in \mathcal{L}^+$, and for $G \in \mathcal{G}$ define $\mu(G) = L(\mathbb{1}_G)$. Notice that $\mathbb{1}_{G \cap H} = \min(\mathbb{1}_G, \mathbb{1}_H) \in \mathcal{L}^+$ and $\mathbb{1}_{G \cup H} = \max(\mathbb{1}_G, \mathbb{1}_H) \in \mathcal{L}^+$ by property (3), so \mathcal{G} is closed wrt finite unions and intersections. By property (4), it is also closed under countable unions.

Furthermore, μ is a nonnegative monotone additive function on \mathcal{G} , with inclusion-exclusion, i.e.

$$\mu(G \cup H) - \mu(G \cap H) = \mu(G) + \mu(H)$$

continuity from below, i.e. for $G_n \uparrow G$

$$\mu(G_n) \uparrow \mu(G)$$

and satisfies $\mu(\Omega) = 1$.

Following ?? (??) and closure of \mathcal{G} under countable union, use μ to construct a (Munroe) outer measure

$$\mu^*(A) = \inf\{\mu(G) : G \in \mathcal{G}, A \subset G\}$$

By ?? (??), μ^* is a countably additive measure on the σ -algebra

$$\mathcal{B} = \{B \subset \Omega : \mu^*(B) + \mu^*(\Omega \setminus B) = 1\}$$

Let μ denote the restriction of μ^* to \mathcal{B} .

Uncertain about above theorem

Should check details of section 1.5 Borgachev

Armed with μ , we now verify that $\mathcal{A} = \sigma(\mathcal{F})$ (the σ -algebra generated by our vector lattice of functions \mathcal{F}) is contained in the domain \mathcal{B} where μ is defined. For $f \in \mathcal{L}^+$, $\{f > c\} \in \mathcal{G}$ for all c because

$$\mathbb{1}_{\{f > c\}} = \lim_n \min(1, n \max(f - c, 0)) \quad (11) \quad \{\text{eq:superlevel-set-in-G}\}$$

Hence $f \in \mathcal{L}^+$ are measurable wrt $\sigma(\mathcal{G})$, but they are also measurable wrt $\sigma(\mathcal{F})$ (since they are monotone limits of things in \mathcal{F}), so $\mathcal{G} \subset \sigma(\mathcal{L}^+) = \sigma(\mathcal{F})$ and by Dynkin π - λ we have $\sigma(\mathcal{G}) = \sigma(\mathcal{F}) = \mathcal{A}$. Thus, it suffices to show $\mathcal{G} \subset \mathcal{B}$.

For $G \in \mathcal{G}$, let $\mathcal{F} \ni f_n \uparrow \mathbb{1}_G$ so

$$\mu^*(G) = \mu(G) = \lim_{n \rightarrow \infty} L(f_n)$$

and since (because μ^* is an outer measure) $\mu^*(G) + \mu^*(\Omega \setminus G) \geq 1$, to show $G \in \mathcal{B}$ it suffices to prove $\mu^*(G) + \mu^*(\Omega \setminus G) \leq 1$ i.e.

$$\mu^*(\Omega \setminus G) \leq \lim_n L(\mathbb{1} - f_n)$$

Let $U_c = \{\mathbb{1} - f_n > c\}$ for $n \in \mathbb{N}$ and $c \in (0, 1)$, so $U_c \supset \Omega \setminus G$ (by monotonicity we must have $(\mathbb{1} - f_n)(x) \equiv 1$ for $x \notin G$) and $\mathbb{1}_{U_c} \leq c^{-1}(\mathbb{1} - f_n)$ by definition. We also have $U_c \in \mathcal{G}$ by ??, so altogether

$$\mu^*(\Omega \setminus G) \leq \mu(U_c) = L(\mathbb{1}_{U_c}) \leq c^{-1}L(\mathbb{1} - f_n)$$

Take $c \rightarrow 1$ and $n \rightarrow \infty$ to get the desired result.

Having defined μ on $\mathcal{A} = \sigma(\mathcal{F})$, it remains to prove $\mathcal{F} \subset \mathcal{L}^1(\mu)$ and that $L(f) = \int_{\Omega} f d\mu$. For $f \in \mathcal{L}^+$ with $f \leq 1$ (bounded), approximate f as the limit of increasing sequence of simple functions

$$f_n = \sum_{j=1}^{2^n-1} j2^{-n} \mathbb{1}_{\{j2^{-n} < f < (j+1)2^{-n}\}}$$

$$L(f_n) = \sum_{j=1}^{2^n-1} j2^{-n} \mu\{j2^{-n} < f < (j+1)2^{-n}\} = \int_{\Omega} f_n d\mu$$

By property (4) and properties of the integral on increasing sequences $\{f_n\}$, taking $n \rightarrow \infty$ yields the desired formula $L(f) = \int_{\Omega} f d\mu$. By considering truncations $\mathcal{L}^+ \ni \min(f, n) \rightarrow f$, which are increasing in n , this extends to non-negative f . By splitting $f \in \mathcal{F}$ as $f = \max(f, 0) - \max(-f, 0)$, this shows that $\mathcal{F} \subset \mathcal{L}^1(\mu)$ with the desired integration formula.

Lastly, the uniqueness of μ follows from Dynkin's π - λ combined with the fact that it is uniquely determined on the class \mathcal{G} , which is closed wrt finite intersections and generates \mathcal{A} as a σ -algebra. \square

===== »»»> Adds lec5

4 Lecture 4: Representation theorems

Recall our current setup:

- \mathcal{F} is a vector lattice of functions over Ω containing constants, i.e. $\mathbb{1} \in \mathcal{F}$
- \mathcal{L}^+ consists of f such that $0 \leq f_n \uparrow f < \infty$ for $f_n \in \mathcal{F}$
- $\mathcal{G} = \{G \subset \Omega : \mathbb{1}_G \in \mathcal{L}^+\}$ are the subsets whose indicators can be realized as monotone limits within \mathcal{F} (i.e. can be well approximated using \mathcal{F} , which we can use L to measure); we use \mathcal{G} as the approximating set when constructing the Munroe outer measure
- For $G \in \mathcal{G}$, define $\mu(G) = L(\mathbb{1}_G) = \lim_n L(f_n)$ and extend using ?? (??) and ?? (??) to the class $\mathfrak{M}_{\mu^*} = \mathcal{B}$ which contains $\mathcal{A} = \sigma(\mathcal{F}) = \sigma(\mathcal{G})$.

Corollary 50

Suppose that in ?? the vector lattice \mathcal{F} is closed wrt uniform convergence. Let $\mathcal{G}_{\mathcal{F}}$ be the class of sets of the form $\{f > 0\}$, $f \geq 0$, $f \in \mathcal{F}$. Then $\mathcal{G}_{\mathcal{F}}$ generates $\mathcal{A} = \sigma(\mathcal{F})$ and we have

$$\begin{aligned} \mu(A) &= \inf\{\mu(G) : A \subset G, G \in \mathcal{G}_{\mathcal{F}}\}, & \forall A \in \mathcal{A} \\ \mu(G) &= \sup\{L(f) : f \in \mathcal{F}, 0 \leq f \leq \mathbb{1}_G\}, & \forall G \in \mathcal{G}_{\mathcal{F}} \end{aligned}$$

Proof. Suffices to verify $\mathcal{G}_{\mathcal{F}}$ equals the \mathcal{G} introduced during the theorem's proof, since we showed $\sigma(\mathcal{G}) = \mathcal{A}$. Taking $c = 0$ in ?? shows $\{f > 0\} \in \mathcal{G}$ for all non-negative $f \in \mathcal{F}$. On the other hand, if $G \in \mathcal{G}$ then $f_n \uparrow \mathbb{1}_G$ for some $f_n \geq 0$ in \mathcal{F} . Letting $f = \sum_{n=1}^{\infty} 2^{-n} f_n$, by uniform convergence of the series we have $f \in \mathcal{F}$. Clearly $f \geq 0$ and $G = \{f > 0\}$. \square

A general fact about vector lattices where signed measures decompose into a positive part and negative part. If ν is a signed measure on Ω , then $\nu = \nu_+ - \nu_-$ for ν_{\pm} unique nonnegative meaasures with disjoint supports. Its total variation decomposes as:

$$\|\nu\| = \nu_+(\Omega) + \nu_-(\Omega)$$

Theorem 51

Let \mathcal{F} be a vector lattice of bounded functions on a set Ω such that $\mathbb{1} \in \mathcal{F}$. Suppose that we are given a linear functional L on \mathcal{F} that is continuous wrt $\|f\| = \sup_{\Omega} |f(x)|$, i.e.

$$\|L\| = \inf\{c : \|L(f)\| \leq c\|f\| \forall f \in \mathcal{F}\} < \infty$$

Then L can be represented as $L = L^+ - L^-$ where $L^+ \geq 0$, $L^- \geq 0$, and for all nonnegative $f \in \mathcal{F}$ we have

$$L^+(f) = \sup_{0 \leq g \leq f} L(g), \quad L^-(f) = - \inf_{0 \leq g \leq f} L(g)$$

In addition, letting $|L| = L^+ + L^-$, we have for $f \geq 0$

$$|L|(f) = \sup_{0 \leq |g| \leq f} |L(g)|, \quad \|L\| = L^+(\mathbb{1}) + L^-(\mathbb{1})$$

Proof. Given two nonnegative $f, g \in \mathcal{F}$ and $h \in \mathcal{F}$ such that $0 \leq h \leq f + g$, can write $h = h_1 + h_2$ where $0 \leq h_1 \leq f$, $0 \leq h_2 \leq g$, $h_1, h_2 \in \mathcal{F}$. Just let $h_1 = \min(f, g)$ and $h_2 = h - h_1$ and verify.

Let L^+ be defined by the previous theorem. We first show additivity on nonnegative functions. For $f, g \in \mathcal{F}$ nonnegative, we have

$$L^+(f + g) = \sup\{L(h) : 0 \leq h \leq f + g\} = \sup\{L(h_1) + L(h_2) : 0 \leq h_1 \leq f, 0 \leq h_2 \leq g\} = L^+(f) + L^+(g)$$

where we used the previous decomposition.

Now we show additivity on arbitrary functions. Let $f = f_1 - f_2$, where f_1, f_2 non-negative. There might be multiple decompositions for the same f , but still

$$L^+(f) = L^+(f_1) - L^+(f_2)$$

since $f_1 + f^- = f_2 + f^+$ and we showed L^+ is additive on nonnegative functions.

Define $L^- = L^+ - L$ and since $L^+(f) \geq L(f)$ for $f \geq 0$ we have that L^- is also nonnegative.

Finally,

$$\begin{aligned} \|L\| &\leq \|L^+\| + \|L^-\| \\ &= L^+(1) + L^-(1) \\ &= 2L^+(1) - L^-(1) \\ &= \sup\{L(2\psi - 1) : 0 \leq \psi \leq 1\} \\ &\leq \sup\{L(h) : -1 \leq h \leq 1\} \\ &\leq \|L\| \end{aligned}$$

□

Corollary 52

Suppose in addition $L(f_n) \rightarrow 0$ for every $f_n \downarrow 0$. Then L^+ and L^- share this property as well, and are defined by nonnegative countably additive measures on $\sigma(\mathcal{F})$ and L has representation

$$L(f) = \int_{\Omega} f d\mu, \quad \forall f \in \mathcal{F}$$

with some signed countably additive measure μ on $\sigma(\mathcal{F})$.

Proof. TODO

□

Here is an analogue of the Riesz representation theorem:

Theorem 53

Let X be a topological space. The formula

$$L(f) = \int_X f d\mu$$

establishes a one-to-one correspondence between Baire measures μ on X and continuous linear functionals L on $\mathcal{C}_b(X)$ with the property

$$\lim_n L(f_n) = 0$$

for every $f_n \downarrow f$.

Proof. Any measure μ on $\mathcal{B}_a(X)$ defines a continuous linear functional on $\mathcal{C}_b(X)$ through the above formula.

Converse follows from ??.

□

See “Banach limit”

Theorem 54 (Dini’s theorem)

On a compact space K , if $\{f_n\} \subset \mathcal{C}(X)$ converges pointwise decreasing to zero, then $\{f_n\}$ converges in the Banach space $\mathcal{C}(X)$ to 0, i.e. converges uniformly to zero.

We get a Riesz representation for compact spaces:

Theorem 55 (Riesz representation theorem)

On a compact Hausdorff space K , every continuous linear functional L on the Banach space $\mathcal{C}(K)$ has

a unique Radon measure μ such that

$$L(f) = \int_K f d\mu, \quad \forall f \in \mathcal{C}(K)$$

Proof. By ??, TODO □

From now, we assume S to be locally compact, second countable, and Hausdorff (lscH). Let $\mathcal{G}, \mathcal{F}, \mathcal{K}$ denote open, closed, and compact sets in S and put $\hat{\mathcal{G}} = \{G \in \mathcal{G}, \bar{G} \in \mathcal{K}\}$. Let $\hat{C}_+ = \hat{C}_+(S)$ denote the class of continuous functions $f : S \rightarrow \mathbb{R}_+$ with compact support (i.e. closure of the set $\{x \in S; f(x) > 0\}$).

We want to extend the idea of invariant (Haar) measure from just groups to more general spaces such as the sphere.

Theorem 56 (Riesz representation)

If S is lscH, then every positive linear functional μ on $\hat{C}_+(S)$ extends uniquely to a measure on S that assigns finite mass to compact sets.

Proof. Kallenberg, “Foundations of modern probability” □

Theorem 57

On every lscH group G there exists, uniquely up to normalization, a left-invariant measure $\lambda \neq 0$ that assigns finite mass to compact sets. If G is compact, then λ is also right-invariant.

Proof. Kallenberg, “Foundations of modern probability” □

Definition 58

Given group G and space S , a *left action* of G on S is a mapping $(g, s) \mapsto gs$ such that $es = s$ and $(gh)s = g(hs)$ for any $g, h \in G$ and $s \in S$, where e denotes the identity element in G .

Similarly, a *right action* is a mapping $(s, g) \mapsto sg$ satisfying similar compatibility conditions.

The action is *transitive* if for all $s, t \in S$ there exists $g \in G$ such that $gs = t$ or $sg = t$ respectively.

All actions are assumed left henceforth.

When G is a topological group, we assume the action is a continuous $G \times S \rightarrow S$ map.

Definition 59

$h : G \rightarrow S$ is *proper* if $h^{-1}K$ is compact in G for any compact $K \subset S$.

If this holds for all $\pi_s(x) = xs$, $s \in S$, we say the group action is proper.

Definition 60

A measure μ on S is G -invariant if $\mu(xB) = \mu B$ for any $x \in G$ and $B \in \mathcal{S}$. This is clearly equivalent to

$$\int f(xs) \mu(ds) = \mu f$$

for any measurable $f : S \rightarrow \mathbb{R}_+$ and $x \in G$.

Theorem 61

If we have lscH group G acting transitively and properly on lscH space S . Then there exists a unique (up to normalization) G -invariant measure $\mu \neq 0$ on S which assigns finite mass to compact sets.

Proof. We first show existence. Fix $p \in S$ and let $\pi = x \mapsto xp : G \rightarrow S$. Letting λ be a left Haar measure on G , define the pushforward $\mu = \lambda \circ \pi^{-1}$ on S . Since π is proper and the Haar measure on G

assigns finite mass to compact sets, μ is a measure on S that assigns finite mass to compact sets. To see G -invariance, for $f \in \hat{C}_+(S)$ and $x \in G$

$$\int_S f(xs)\mu(ds) = \int_G f(xyp)\lambda(dy) = \int_G f(y)p\lambda(dy) = \mu f$$

by invariance of λ .

Now we consider uniqueness. Let μ be arbitrary G -invariant measure on S assigning finite mass to compact sets. Define the subgroup

$$K = \{x \in G : xp = p\} = \pi^{-1}\{p\}$$

(the stabilizer of p , subgroup leaving p fixed) and note K is compact (since π is proper). Let ν be the normalized Haar measure on K , and define

$$\bar{f}(x) = \int_K f(xk)\nu(dk), \quad x \in G, f \in \hat{C}_+(G)$$

At each point x , \bar{f} takes f and “smooths things out” using K translated to x .

If $xp = yp$ then $y^{-1}xp = p$ and so $y^{-1}x =: h \in K$ which implies $x = yh$. Hence, left invariance of ν yields

$$\bar{f}(x) = \bar{f}(yh) = \int_K f(yhk)\nu(dk) = \int_K f(yk)\nu(dk) = \bar{f}(y)$$

So the mapping $f \mapsto f^*$ given by

$$f^*(s) = \bar{f}(\pi^{-1}\{s\}) \equiv \bar{f}(x), \quad s = xp \in S, x \in G, f \in \hat{C}_+(G)$$

is well defined, and for any $B \subset (0, \infty)$ we have

$$(f^*)^{-1}B = \pi(\bar{f}^{-1}B) \subset \pi[(\text{supp } f) \cdot K]$$

where $(\text{supp } f) \cdot K$ is the support of f “convolved with K ” via the group action. Hence, the RHS is compact (both $\text{supp } f$ and K compact) and since π and the action are continuous. Therefore f^* has compact support.

Also, \bar{f} is continuous (by group operation cts and dominated convergence), so $\bar{f}^{-1}[t, \infty)$ is closed and hence compact for every $t > 0$.

??? So f^* is something we can integrate against μ .

We may now define functional λ on $\hat{C}_+(G)$ by $\lambda f = \mu f^*$ for $f \in \hat{C}_+(G)$. Linearity and positivity of λ are clear from the corresponding properties of the mapping $f \mapsto f^*$ and the measure μ . We note that λ is finite on $\hat{C}_+(G)$ since μ is locally finite, so by Theorem 6 we can extend λ to a measure on G that assigns finite mass to compact sets.

To see λ left invariant, for $f \in \hat{C}_+(G)$ and define $f_y(x) = f(yx)$. Then for $s = xp \in S$ and $y \in G$ we have

$$f_y^*(s) = \bar{f}_y(x) = \int_K \bar{f}_y(xk)\nu(dk) = \bar{f}(yx) = f^*(ys)$$

Hence by invariance of μ we have

$$\int_G f(yx)\lambda(dx) = \lambda f_y = \mu f_y^* = \int_S f^*(ys)\mu(ds) = \mu f^* = \lambda f$$

So λ is the Haar measure.

Now fix $g \in \hat{C}_+(S)$ and put $f(x) = g(xp) = g \circ \pi(x)$ for $x \in G$. Then $f \in \hat{C}_+(G)$ because $\{f > 0\} \subset \pi^{-1}\text{supp } g$ which is compact since π is proper. By definition of K , for $s = xp \in S$ we have

$$f^*(s) = \bar{f}(x) = \int_K f(xk)\nu(dk) = \int_K g(xkp)\nu(dk) = \int_K g(xp)\nu(dk) = g(s)$$

so we’ve found an inverse for the $*$ operation, so

$$\mu g = \mu f^* = \lambda f = \lambda(g \circ \pi) = (\lambda \circ \pi^{-1})g$$

which shows $\mu = \lambda \circ \pi^{-1}$. Since λ is unique up to normalization, the same thing is true for μ . \square

5 Lecture 5: Extreme point representation of measures

2020-02-04

Throughout, let E be a real TVS.

Definition 62

For convex $A \subset E$, an *open segment* is a subset of type

$$\{(1 - \lambda)a + \lambda b : \lambda \in (0, 1)\}, \quad a \neq b \in A$$

$x_0 \in A$ is an *extreme point* of A if it belongs to no open segment of A , i.e.

$$\begin{aligned} &(\exists \lambda \in [0, 1] : x_0 = (1 - \lambda)a + \lambda b) \\ &\Rightarrow x_0 = a \text{ or } x_0 = b \end{aligned}$$

Closed $B \subset A$ is an *extreme subset* of A if

$$\begin{aligned} &(\exists \lambda \in (0, 1) : (1 - \lambda)a + \lambda b \in B) \\ &\Rightarrow \{a, b\} \subset B \end{aligned}$$

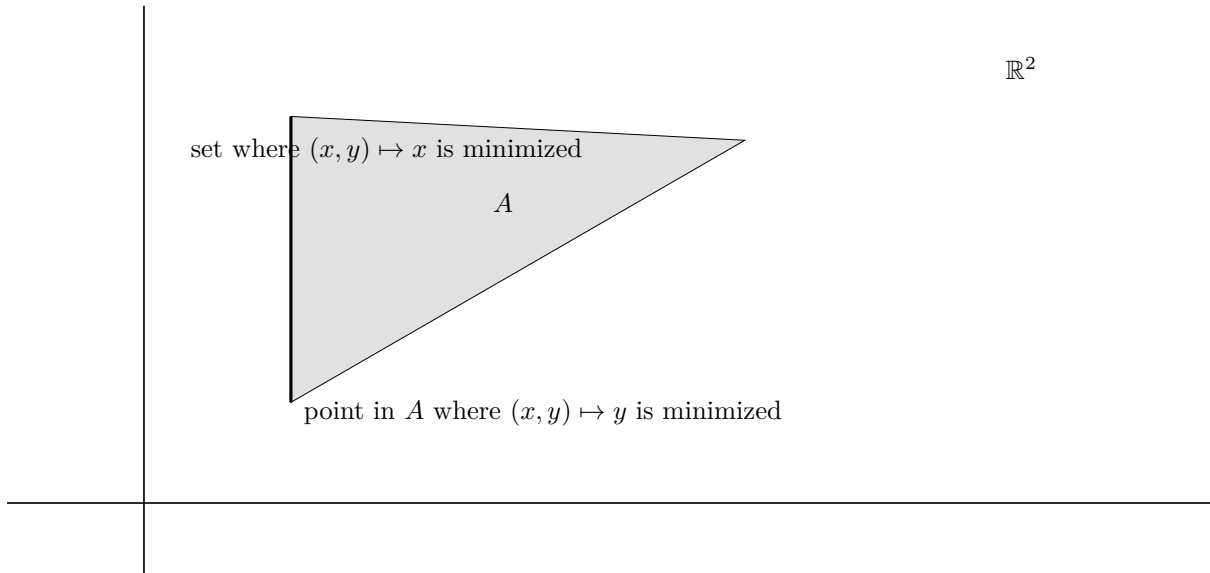


Figure 3: On a triangle, the extreme points are the vertices and the extreme subsets are the sides

Lemma 63

Let E real Hausdorff TVS, $A \subset E$ nonempty compact convex, f continuous linear functional on E , $\beta = \inf_{x \in A} f(x)$. Then $B = A \cap f^{-1}(\beta)$ is nonempty, compact, extreme subset of A .

Proof. β exists because A is compact.

Continuity and linearity of f ensure B is closed and convex. Check nonempty and compact.

To show B extreme, suppose $(1 - \lambda)a + \lambda b \in B$ for $a, b \in A$ and $\lambda \in (0, 1)$. If, for example, $a \notin B$, then $f(a) > \beta$ by definition of B and so by linearity

$$f((1 - \lambda)a + \lambda b) = (1 - \lambda)f(a) + \lambda f(b) > (1 - \lambda)\beta + \lambda\beta = \beta$$

contradicting $(1 - \lambda)a + \lambda b \in B$

TODO: finish

□

Theorem 64

¹ E real Hausdorff LCTVS, A nonempty compact convex subset of E , then A is the closed convex hull in

E of the set of extreme points of A .

This is saying something along the same lines as Caratheodory's theorem for convex things.

Proof. We first show each nonempty extreme subset $X \subset A$ contains an extreme point of A . Let \mathfrak{X} consist of extreme subsets of A contained in X . \mathfrak{X} is nonempty (by Lemma 13), so partially order \mathfrak{X} by inclusion. Notice the intersection of any chain is a non-empty compact set $\in \mathfrak{X}$ because it is the intersection of nonempty compact sets (c.f. Hausdorff's theorem), hence by Zorn's lemma \mathfrak{X} possesses a minimal element, say Y .

It remains to show Y is a singleton. Otherwise, Y would contain $x \neq y$ and since E is Hausdorff and LC, (by Hahn-Banach separating hyperplane version) there exists continuous linear function f on E such that $f(x) < f(y)$. By Lemma 13, $Z = Y \cap f^{-1}(\inf f(Y))$ is a nonempty extreme subset of A that does not contain y . Thus $Z \subsetneq Y$, contradicting minimality of Y .

In the second step, let B be the closed convex hull in E of the set of all extreme points of A . B is compact, convex, and contained in A . To show $B = A$, it suffices to show $A \subset B$ is empty. Suppose towards contradiction $x_0 \in A \setminus B$, then by Hahn-Banach theorem there exists (separating) continuous linear functional f on E such that $\inf_B f(x) > f(x_0)$. Then by Lemma 13 $W = A \cap f^{-1}(\inf_A f(x))$ is a nonempty extreme subset of A disjoint from B . However, by the previous part W would contain an extreme point of A , which is a contradiction since $W \cap B = \emptyset$. \square

Proposition 65

Suppose E is Hausdorff real LCTVS, $K \subset E$ compact whose closed convex hull A is compact. Then each extreme point of A belongs to K

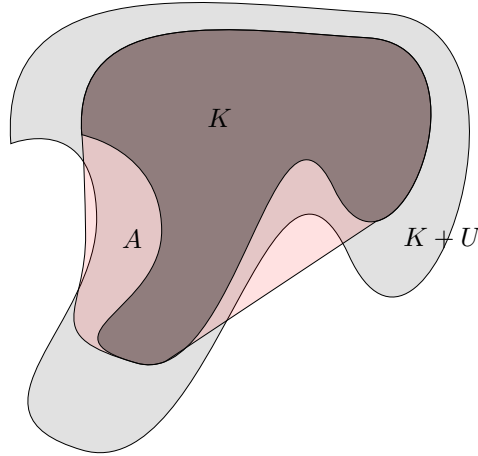


Figure 4: cvx-hull-contains-extreme-pts

Proof. Let $x \in A$ be extreme point. For any closed convex nbd $0 \in U \subset E$, by compactness of K there exists finitely many $a_i \in K$ such that $a_i + U$ cover K . Let $A_i = \overline{\text{conv}(K \cap (a_i + U))}$, each A_i is compact because $A_i \subset K$ and K is compact. Then $\text{conv}(\cup_i^n A_i)$ is compact and $K \subset \text{conv}(\cup_i^n A_i) \subset A$, so we must have $A = \text{conv}(\cup_i^n A_i)$.

Hence, $x = \sum_i^n \lambda_i x_i$ with $x_i \in A_i$, $\lambda_i \geq 0$, $\sum_i \lambda_i = 1$. As $x \in A$ is extreme point, x must coincide with some x_i . Thus, $x \in A_i \subset a_i + U$, so $x \in K + U$. Since K is closed and U is an arbitrary nbd of 0, $x \in K$ as desired. \square

Example 66

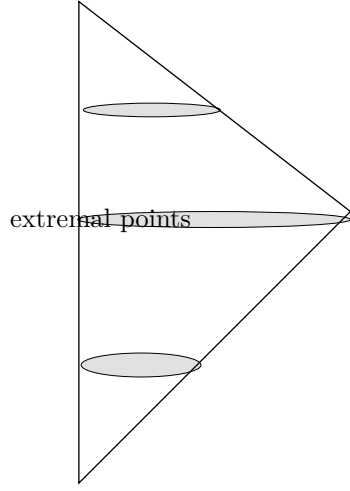


Figure 5: higher-dim-extreme-pt

A compact convex set A need not be the convex hull of its extreme points. Take $E = \ell^\infty$, $e_n = \delta_n \in E$, A the closed convex hull in E of e_n/n for $n \in \mathbb{N}$. By (TODO: prop), the extreme points of A are 0 and the points $\{e_n/n\}$. A is compact and contains all points $x = \sum_{n \geq 1} \lambda_n e_n/n$, λ_n a convex combinations. TODO: finish

Definition 67

For a Banach space $(X, \|\cdot\|_X)$, let X' be the set of bounded linear functionals $L : X \rightarrow \mathbb{R}$, i.e. $\sup_{x \in X} |L(x)|/\|x\| \leq c$ for some $c \geq 0 \Leftrightarrow L$ is linear and continuous.

Given $L \in X'$, write $\|L\|_{X'}$ for smallest c that works $\|L\|_{X'} = \inf\{|L(x)| : \|x\| = 1\}$.

FACT: $(X', \|\cdot\|_{X'})$ is a Banach space. Each $x \in X$ defines a linear map $X' \rightarrow \mathbb{R}$ via evaluation

$$e_x = L \mapsto L(x)$$

The *weak-* topology* on X' is the weakest/coarsest (i.e. initial topology) $\tau(X', \{e_x\}_{x \in X})$ that makes all of these maps continuous.

Example 68 (Riesz representation theorem)

Let S be a compact Hausdorff space, $X = \mathcal{C}(S)$ continuous functions from S to \mathbb{R} . X is a Banach space with the sup-norm $\|x\| = \sup_{s \in S} |x(s)|$. Then X' finite signed measures on S . associated with any finite signed measure μ . is the continuous linear functional $L(x) = \int_S x(s) \mu(ds)$.

We want to know what is $\|\cdot\|_{X'} = \|\cdot\|_{\mu(S)}$? If $L(x) = \int_S x(s) \mu(ds)$ then by definition

$$\begin{aligned} \|L\|_{X'} &= \sup\{|L(x)| : \|x\|_X = 1\} \\ &= \sup\left\{\left|\int_S x(s) \mu(ds)\right| : \sup_{s \in S} |x(s)| = 1\right\} \\ &= \sup\left\{\left|\int_S x(s) \mu^+(ds) - \int_S x(s) \mu^-(ds)\right| : -1 \leq |x(s)| \leq 1\right\} \end{aligned}$$

We know μ^+ and μ^- are perpendicular; their supports are disjoint, so

$$\|L\|_{X'} = \sup\left\{\left|\int_{S^+} x(s) \mu^+(ds) - \int_{S^-} x(s) \mu^-(ds)\right| : -1 \leq |x(s)| \leq 1\right\}$$

Take $x = \mathbb{1}_{S^+} - \mathbb{1}_{S^-}$ to conclude $\|L\|_{X'} = \mu^+(S^+) + \mu^-(S^-) = \mu^+(S) + \mu^-(S) = |\mu|(S) = \|\mu\|_{TV}$. Hence, $(\mathcal{C}(S), \|\cdot\|_\infty)$ has dual $(M(S), \|\cdot\|_{TV})$ and the weak-* topology on $M(S)$ is the weakest/coarsest/smallest

topology that makes continuous all maps $\mu \mapsto \int_S x(s)\mu(ds)$ for $x \in C(S)$. Notice that this is strictly weaker than the TV topology, because for example two unit point masses have TV norm 2 but $\int x(s)\delta_{s'}(ds) = x(s') \approx x(s'') = \int x(s)\delta_{s''}(ds)$ when $s' \approx s''$.

This is the story for compact spaces. What about for only locally compact spaces?

Let T Hausdorff LC, $M(T)$ real bounded signed Radon measures on T .

Definition 69

$C_0(T)$ are the continuous functions vanishing at infinity, i.e. $f \in C(T)$ such that $\lim_{x \rightarrow \infty} f(\pm x) = 0$.

View $M(T)$ as the (Banach) dual of $C_0(T)$ equipped with weak-* topology. The set $M_+^1(T)$ of positive measures in $M(T)$ having total mass at most one is compact and convex, because:

Theorem 70 (Banach-Alaoglu)

The unit ball in $(X', \|\cdot\|_{X'})$ is compact in the weak-* topology.

We will show that the extreme points of $M_+^1(T)$ are 0 and the Dirac measures δ_t for $t \in T$.

It is clear that 0 is an extreme point of $M_+^1(T)$.

Suppose $\mu \neq 0$ is another element in $M_+^1(T)$, then it suffices show $K = \text{supp } \mu$ is a single point. If $t_1 \neq t_2 \in K$, then by Hausdorffness choose $U_1 \ni t_1$ and $U_2 \ni t_2$ disjoint. Then $m = \mu(U_1)$ satisfies $0 < m < 1$.

Define two measures $\alpha = (\mu|_{U_1})/m$ (μ restricted and renormalized) and $\beta = (\mu - m\alpha)/(1 - m)$ what's left over after subtracting $m\alpha$. Then $\alpha, \beta \in M_+^1(T)$, $\alpha \neq \beta$, and $\mu = m\alpha + (1 - m)\beta$, contradicting μ extremal. Hence, K is a single point.

A similar argument shows that if T is compact, then the set of positive measures of unit total mass is compact and convex, and that its extreme points are the Dirac measures δ_t .

Theorem 71 (Stone-Weierstrass)

Let E be a subalgebra of $C(S)$. Suppose E separates points and contains constants, then E is dense in $C(S)$.